

## COMPARAISONS DIACHRONIQUES ET SUBSTANCIALISATION DES VARIABLES : EXEMPLE DE L'ÉVOLUTION DES INÉGALITÉS SCOLAIRES

Guy MOREL<sup>1</sup>

**RÉSUMÉ** – *Mesurer l'évolution de l'inégalité scolaire est un problème difficile. Il faut construire un indice d'inégalité mais aussi prendre en compte les changements sociaux qui affectent les catégories sociales et la valeur des diplômes. La démocratisation « quantitative » s'impose quel que soit l'outil statistique utilisé, il n'en est pas de même de la démocratisation « qualitative ». Nous proposons de l'analyser en construisant une variable latente « potentiel scolaire » qui élimine la démocratisation « quantitative » en donnant une valeur relative aux diplômes. L'étude de 19 générations triennales entre 1920 et 1976 montre que la démocratisation « qualitative » porte principalement sur les enfants des catégories agricoles. Ce résultat est semblable à celui obtenu avec le modèle log-multiplicatif de variation uniforme. Peut-on parler globalement d'une démocratisation « qualitative » ?*

**MOTS-CLÉS** – Démocratisation, Inégalités de scolarisation, Modèle log-multiplicatif, Variable latente

**SUMMARY** – Diachronic comparisons and substancialisation of variables. An example: the evolution of educational inequalities

*The evolution of educational inequalities is difficult to measure. It is necessary to build an indicator of inequality but also to take into account the social changes which affect the social classes and the value of the diplomas. Whichever the statistical tool used, there is always "quantitative" democratization. It is not the same with "qualitative" democratization. We propose to analyze it by building a latent variable called "educational potential". We eliminate "quantitative" democratization by giving a relative value to the diplomas. The study of 19 triennial generations between 1920 and 1976 shows that "qualitative" democratization relates mainly to the children of the agricultural categories. This result is similar to that obtained with the log-multiplicative layer effect model. On the whole, can we speak about a "qualitative" democratization?*

**KEY-WORDS** – Democratization, Educational level inequalities, Latent variable, Log-multiplicative model

La comparaison diachronique la plus simple est l'étude d'un processus temporel saisi à deux périodes différentes. Prenons l'exemple de la comparaison des taux de réussite au baccalauréat des enfants d'agriculteurs nés dans les années 1947-1949 et de ceux nés dans les années 1974-1976. Les deux pourcentages valent respectivement 19.2 % et 71.6 %. Ils donnent à voir une forte augmentation de l'accès au baccalauréat des enfants d'agriculteurs. On peut cependant se poser la question de la légitimité de la

---

<sup>1</sup> Laboratoire de Mathématiques et Physique Théorique (LMPT), UMR 6083 et CITERES Equipe Villes Sociétés Territoires, UMR 6173, Département de sociologie, Université François Rabelais, 3 rue des Tanneurs, 37041 Tours Cedex, morel@univ-tours.fr

comparaison entre la variable réussite au baccalauréat définie sur la première cohorte (1947-1949) et celle définie sur la deuxième cohorte (1974-1976). D'une part, on peut se demander s'il est intéressant de comparer les deux populations concernées. Elles portent bien le même nom : enfants d'agriculteurs, mais la catégorie socioprofessionnelle des agriculteurs est celle qui s'est le plus métamorphosée en quantité et en contenu entre 1947 et 1976. D'autre part, on peut se demander si la variable réussite au baccalauréat mesure quelque chose de semblable sur ces deux cohortes. L'évolution du système scolaire ne permet pas de répondre par l'affirmative. Le seuil baccalauréat ne représente plus la même valeur scolaire, la même potentialité sociale. On se retrouve devant un problème semblable à celui de la définition du seuil de pauvreté dans des pays très différents.

En sciences humaines, les comparaisons dans le temps ou dans l'espace posent presque systématiquement ce type de questions [M. Duru-Bellat, A. Kieffer, 2000]. Les mesures et catégorisations peuvent y être fortement instables, la substantialisation des variables est un risque courant. La comparaison de simples pourcentages se heurte à des difficultés spécifiques que l'on ne retrouve pas, par exemple, lors de la comparaison des pourcentages de réussite dans la traversée d'un labyrinthe de deux souches de souris. L'épreuve est standardisée, son environnement est fixé avec beaucoup de détails et chacune des souches de souris est construite pour représenter une diversité stable dans le temps. Si on compare les performances à cette épreuve d'une même souche à dix ans d'intervalle, on espère retrouver la même chose, c'est une répétition, une vérification. Une hypothèse de différence de performances sera liée à une modification bien spécifiée de la population ou de la mesure. On peut modifier l'épreuve, changer de souche de souris ou comparer des souris de même souche ayant eu des élevages différents. Ceci est bien différent des performances au baccalauréat à dix ans d'intervalle de la "souche" enfants d'agriculteurs. L'évolution des pourcentages peut être reliée à trois types de modifications sociales : celles des caractéristiques de la catégorie agriculteurs, celles des caractéristiques du niveau baccalauréat, mais aussi à des modifications plus générales qui ne s'épuisent pas dans leurs conséquences sur les caractéristiques précédentes. Bien entendu cette typologie n'est pas vraiment opératoire, les évolutions sociales sont le plus souvent interdépendantes. C'est justement cette interdépendance qui est simplifiée en substantialisant la variable réussite au baccalauréat des enfants d'agriculteurs. L'interprétation de la comparaison ne prend pas en compte une éventuelle modification du positionnement social de la catégorie ou de la mesure étudiée.

Lorsque la comparaison diachronique prend en compte l'évolution de plusieurs populations et/ou de plusieurs mesures une autre difficulté apparaît. Le chercheur doit choisir les résumés statistiques dont il étudiera l'évolution afin de prendre position par rapport à ses hypothèses. Cette difficulté n'est pas spécifique aux sciences humaines même si le choix des indices dépend des hypothèses du chercheur. Dans le long débat autour de l'évolution comparée des inégalités scolaires, initié par J.-C. Combessie [1984], c'est principalement cette dernière difficulté qui est traitée. Nous allons partir de ce débat et regarder ses suites sur la manière de traiter ce type de problème, plus précisément l'évolution des taux de réussite à différents niveaux scolaires pour plusieurs origines sociales. Les niveaux scolaires sélectionnés sont généralement considérés comme des indicateurs stables au cours du temps. Dans la deuxième partie nous essaierons de relativiser leur valeur sociale en prenant en compte leur niveau de diffusion dans chaque cohorte. Nous appliquerons ensuite ce point de vue à un exemple qui porte sur 19 générations triennales (entre 1920 et 1976), 11 origines sociales et 7 niveaux de diplôme. Dans la dernière partie, nous reviendrons, à partir de cet exemple,

sur la traduction par un indice statistique de la notion de « démocratisation du système scolaire ». Nous verrons comment l'interprétation de ces indices globaux peut être biaisée si l'on ne fait pas intervenir une connaissance plus fine des données. La séparation entre démocratisation *quantitative* et démocratisation *qualitative*<sup>2</sup> [A. Prost, 1986, p. 12] pose aussi problème, elle peut être modélisée de différentes façons.

## 1. UN DÉBAT ET APRÈS

Comme point de départ, nous allons nous servir du débat autour de l'évolution comparée des inégalités, initié par J.-C. Combessie [1984]. Depuis 1984, au moins une vingtaine d'articles l'ont enrichi. Ils ont été publiés principalement par la *Revue française de sociologie* et par *Mathématiques et Sciences Humaines*.

À la base de ce débat, il y a un exemple déjà utilisé en 1973 par Raymond Boudon dans son livre *L'inégalité des chances*. Il donne pour deux tranches d'âge et deux catégories sociales extrêmes les taux de scolarisation dans le secondaire long des enfants anglais nés dans la première moitié du siècle.

C.S.P. du père	Date de naissance	
	Avant 1910	Entre 1935 et 1940
Professions libérales, cadres et personnel de direction	37 %	62 %
Ouvriers semi qualifiés et non qualifiés	1 %	10 %

TABLEAU 1. Taux de scolarisation

Ce tableau a été construit à partir d'un rapport de J. Westergaard et A. Little [1967]. Il sera analysé comme s'il était issu de données exhaustives ou d'un échantillon suffisamment grand pour que les fluctuations d'échantillonnage aient de grandes chances d'avoir des effets négligeables sur l'interprétation sociologique. Cette position est adoptée dans la plupart des articles prenant part au débat. Pour la critique des sources statistiques ayant conduit à l'exemple traité nous renvoyons à l'article de J.-P. Grémy [1984].

Dans cet exemple les comparaisons deux à deux des pourcentages ne posent pas de problème. Les écarts paraissent tous importants, il y a inégalité de réussite entre les deux origines sociales pour les deux périodes considérées et il y a une évolution positive des taux de scolarisation entre les deux périodes pour chacune des catégories. Les difficultés commencent quand on se pose la question d'une diminution ou d'une augmentation de l'inégalité des chances entre ces deux périodes. Suivant l'indice statistique utilisé pour caractériser l'inégalité de réussite entre les deux catégories, on n'obtient pas toujours la même réponse.

Si cet exemple est souvent repris dans des livres de méthodologie ou de statistique c'est que les points de vue privilégiant écarts absolus et écarts relatifs se contredisent quand on compare l'évolution des taux de scolarisation dans le secondaire long des deux catégories sociales. Lorsqu'on prend la différence des pourcentages comme indice d'inégalité entre les enfants de cadres et les enfants d'ouvriers, on observe une

<sup>2</sup> La démocratisation « quantitative » correspond à l'augmentation du niveau d'instruction général, la démocratisation « qualitative » renvoie elle à l'amélioration de l'égalité des chances.

augmentation de l'inégalité ( $62\% - 10\% > 37\% - 1\%$ ). Si on prend, au lieu de la différence, le rapport des pourcentages comme indice d'inégalité entre les enfants de cadres et les enfants d'ouvriers, on observe une diminution de l'inégalité. Le taux de scolarisation des enfants de cadres est égal à 37 fois celui des enfants d'ouvriers dans la période avant 1910, alors que le rapport des taux n'est plus que de 6.2 ( $62/10$ ) entre 1935 et 1940.

Dans la première partie de son article, J.-C. Combessie [1984, p. 235] passe en revue les opérations élémentaires que l'on peut faire sur ces quatre pourcentages pour porter un jugement sur l'évolution des inégalités. Il en décrit 24, plusieurs étant équivalentes. D. Merllié [1985, p. 642] montre qu'il y a trois manières fondamentales de procéder, les deux décrites précédemment (différence et rapport) et le rapport des taux de non scolarisation dans l'enseignement secondaire long. Ces taux de non scolarisation sont donnés dans le tableau suivant.

C.S.P. du père	Date de naissance	
	Avant 1910	Entre 1935 et 1940
Professions libérales, cadres et personnel de direction	( $100\% - 37\%$ ) 63 %	( $100\% - 62\%$ ) 38 %
Ouvriers semi qualifiés et non qualifiés	( $100\% - 1\%$ ) 99 %	( $100\% - 10\%$ ) 90 %

TABLEAU 2. Taux de non scolarisation

Le point de vue des écarts absolus donne obligatoirement les mêmes conclusions qu'avec les taux de scolarisation ( $90\% - 38\% = 62\% - 10\% > 99\% - 63\% = 37\% - 1\%$ ). Il n'en est pas de même avec le point de vue des écarts relatifs, d'ailleurs dans cet exemple il conduit à des résultats opposés. Le rapport des taux de non scolarisation indique une augmentation de l'inégalité alors que les rapports des taux de scolarisation nous ont conduit à un jugement différent. Avant 1910, le taux de non scolarisation des enfants d'ouvriers est égal à  $99/63 = 1.57$  fois celui des enfants de cadres, pour la période 1935-1940 ce rapport est beaucoup plus grand :  $90/38 = 2.37$ .

Cette structure des concordances et des oppositions entre les conclusions produites par ces trois regards n'est pas constante [D. Merllié, 1985, p. 639], elle dépend des données. La contradiction possible à l'intérieur du point de vue des écarts relatifs est particulièrement choquante, elle conduit J. Prévot [1985, p. 605] à imposer le principe de dualité :

*il paraît naturel et raisonnable de soumettre à ce principe tout indice relatif à l'évolution de proportions dichotomiques (ici taux d'admis et taux d'exclus), à savoir qu'appliqué à l'évolution des proportions complémentaires (les taux d'exclus) cet indice doit conduire à la « même mesure » (i.e. même sens et même intensité) que lorsqu'on l'applique aux proportions premières (les taux d'admis).*

Le point de vue des écarts absolus vérifie ce principe de dualité. Il n'est pas le seul, un autre regard très utilisé est de considérer les rapports des chances et de faire des comparaisons relatives. Sur notre exemple nous allons considérer le rapport des chances d'être scolarisé dans le secondaire long plutôt que non scolarisé.

C.S.P. du père	Date de naissance	
	Avant 1910	Entre 1935 et 1940
Professions libérales, cadres et personnel de direction	(37 % / 63 %) 0.59	(62 % / 38 %) 1.63
Ouvriers semi qualifiés et non qualifiés	(1 % / 99 %) 0.01	(10 % / 90 %) 0.11

TABLEAU 3. Rapports des chances

Avant 1910, le rapport des chances des enfants de cadres (37/63) est égal à 58.14 fois le rapport des chances des enfants d'ouvriers (1/99), pour la période 1935-1940 ce rapport de rapports des chances est plus petit : 14.68. L'inégalité semble donc avoir diminué avec ce nouvel éclairage.

Le principe de dualité énoncé plus haut semble<sup>3</sup> éliminer les indices d'inégalité basés sur le rapport des taux de scolarisation ou celui des taux de non scolarisation. Parmi les deux points de vue restant, L.-A. Vallet [1988] défend l'utilisation du rapport des chances plutôt que celle de la différence des pourcentages. Le rapport des chances a de bonnes propriétés mathématiques<sup>4</sup> et il se prête bien à une interprétation sociologique. Ceci explique le succès de la régression logistique et des modèles log-linéaires pour traiter de l'évolution d'un ensemble de pourcentages. Les autres points de vue continuent cependant d'être utilisés et les problèmes que nous venons de soulever, à partir d'un débat lancé depuis plus de vingt ans, sont encore souvent oubliés. On peut parfois se demander si le point de vue choisi n'est pas celui qui va dans le sens de l'hypothèse préférée. L'exemple donné (p. 64) est tiré d'un rapport destiné au président de la république et rédigé par la Cour des comptes [2003, p. 220]. Il a l'avantage de ne pas avoir d'auteur référencé.

Le graphique montre une extraordinaire constance, au cours du temps, de la différence des pourcentages d'accès à l'enseignement supérieur des enfants de cadres et des enfants d'ouvriers, environ 40 %. Le point de vue des écarts absolus conduit à une constance de cette inégalité d'accès à l'enseignement supérieur. Le point de vue des écarts relatifs conduit lui à la conclusion énoncée dans le titre : une relative démocratisation des formations supérieures. On peut d'ailleurs prédire que ce type de démocratisation va continuer car les enfants de cadres n'ont plus beaucoup de potentialité de progression ( $100/77 = 1.3$ ) alors que les enfants d'ouvriers ont une belle croissance à venir ( $100/35 = 2.9$ ). L'auteur garde l'échelle des pourcentages pour son graphique, s'il l'avait fait en utilisant les rapports des taux d'accès aux formations supérieures sa conclusion ne nous aurait posé aucun problème. Il ne se base cependant pas uniquement sur les écarts relatifs. Bien que ces écarts soient très différents, multiplication par 3.6 pour les enfants d'ouvriers et seulement 1.6 pour les enfants de cadres, il conclut simplement à « une relative démocratisation des formations supérieures ». Ce qualificatif « relative » ne doit pas renvoyer au « relatif » des écarts considérés mais aux fortes disparités en termes d'origine sociale que l'auteur constate dans la population étudiante actuelle. Le titre me semble tout de même trompeur.

<sup>3</sup> L'introduction d'une variable latente donne un point de vue qui conduit à un principe de dualité moins fort [G. Morel, 2003, p. 19]. Les rapports de pourcentages n'y sont pas disqualifiés.

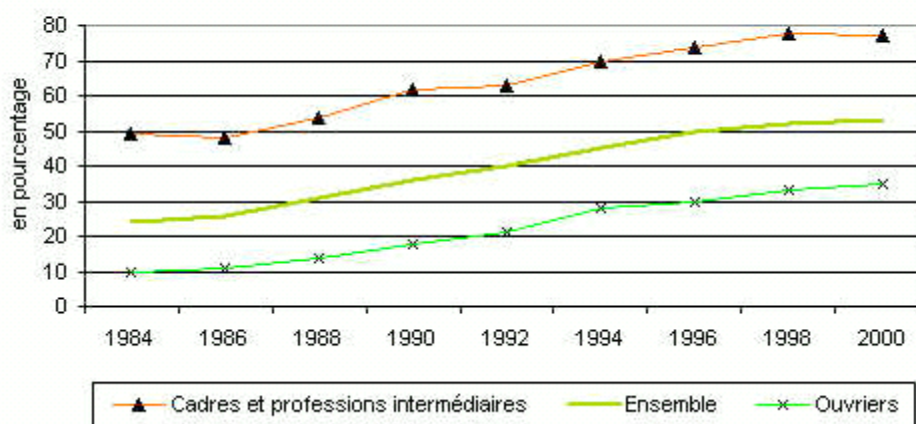
<sup>4</sup> Cf. par exemple H. Léridon et L. Toulemon [1997, p. 247].

*La gestion du système éducatif*, Paris, Cour des comptes, 2003, p. 220.

## 2 – Une relative démocratisation des formations supérieures

Ce phénomène de massification s'est accompagné d'une relative démocratisation de l'enseignement supérieur et le fort développement des scolarités secondaires, intervenu entre 1985 et 1995, a progressivement réduit les inégalités sociales. Si l'écart reste important, puisque aujourd'hui 77 % des enfants de cadres accèdent à l'enseignement supérieur, contre seulement 35 % des enfants d'ouvriers, on constate cependant que les chances d'entrée à l'université se sont davantage élevées pour les couches populaires que pour les couches favorisées de la population. Entre 1985 et 2000, les possibilités de suivre des études supérieures, parmi les jeunes de 20 et 21 ans, ont ainsi été multipliées par 3,6 pour les enfants d'ouvriers, pour 2,2 en moyenne et 1,6 pour les enfants de cadres.

**Accès à l'enseignement supérieur des jeunes de 20/21 ans  
selon leur origine sociale, de 1984 à 2000**



Source : ministère de l'éducation nationale

Cependant la composition de la population étudiante en termes d'origine sociale reste marquée par de fortes disparités et les catégories les plus favorisées demeurent largement sur-représentées au détriment des catégories plus modestes. Quelle que soit la filière considérée, à l'exception des sections de techniciens supérieurs (STS), près du tiers des étudiants ont des parents cadres supérieurs ou exerçant une profession libérale, proportion nettement supérieure à celle constatée auprès de l'ensemble des jeunes. Le phénomène est encore plus marqué dans les classes préparatoires aux grandes écoles (CPGE) et dans les disciplines de santé, où respectivement 50 % et 44 % des étudiants sont issus de cette catégorie sociale. En revanche les filières technologiques courtes (IUT et STS) recrutent davantage parmi les enfants d'ouvriers et d'employés, puisque ceux-ci représentent respectivement 34 % des inscrits en IUT et 42 % de l'effectif des STS.

J.-C. Combessie [2004], l'initiateur du débat en 1984, donne plusieurs exemples de cette attirance vers ce qu'il appelle la conclusion « rose », celle de la diminution des inégalités. Ce comportement l'étonne d'autant plus qu'il constate que de nombreux chercheurs de ce champ font leur le constat suivant : « aucun modèle, aucune mesure, ne sont la meilleure ni même la bonne façon de représenter l'inégalité des chances à l'école »<sup>5</sup> [R. Mare, 1981]. Plutôt que de laisser le chercheur choisir sa mesure de l'inégalité il propose d'en utiliser systématiquement trois ou quatre et d'indiquer le nombre de conclusions donnant le même diagnostic d'inégalité. Nous voyons que la construction d'une mesure de l'inégalité à partir de deux pourcentages continue de poser problème vingt ans après le début du débat. Aucun argument mathématique ne peut être définitivement convaincant. C'est un problème de choix d'une échelle de mesure [G. Morel, 2003, p. 7].

Lorsqu'on prend en compte les taux de scolarisation à différents niveaux scolaires pour les enfants de plusieurs origines sociales, le choix d'un indice d'inégalité devient encore plus problématique. Lors du débat de nombreux indices ont été proposés, ils font principalement référence à trois classes d'indices. Celle basée sur les  $\varphi$ -divergences qui englobent la distance du khi-deux [J.-P. Florens, 1984] ; la classe des indices construits à partir des courbes de concentration de Gini [M. Barbut, 1984 ; É. Terouanne, 1995] ; enfin celle qui repose sur la modélisation log-linéaire qui traduit la constance des inégalités par une absence d'interaction [L.-A. Vallet, 1988].

Les indices de contingence construits à partir de la distance du khi-deux sont encore parfois utilisés. Mais le cadre le plus courant est celui de la modélisation log-linéaire<sup>6</sup>. Le degré d'association entre origine sociale et niveau scolaire ainsi que son évolution au cours des générations se trouvent résumés dans les interactions d'ordre trois ( $\lambda_{ijk}$ ) et celles d'ordre deux entre origine sociale et niveau scolaire ( $\lambda_{ij}$ ) [Y. Xie, 1992, p. 381]. Lorsque ces interactions sont nulles, les variables *origine sociale* et *niveau scolaire* sont indépendantes dans chaque cohorte. L'indépendance traduit naturellement l'égalité parfaite : la répartition des niveaux scolaires est la même pour toutes les origines sociales dans une même cohorte. Pour étudier l'évolution des inégalités les chercheurs modélisent l'interaction complète entre origine sociale et niveau scolaire ( $\lambda_{ij} + \lambda_{ijk}$ ) en utilisant un paramètre indexé par les cohortes, dans un modèle non saturé. L. Goodman et M. Hout [1998] proposent une approche sous forme de régression :  $\lambda_{ij} + \lambda_{ijk} = \lambda_{ij} + \psi_{ij}\phi_k$ . Elle contient un grand nombre de modèles, dont le plus utilisé dans l'évolution des inégalités scolaires est le modèle log-multiplicatif introduit par Yu Xie [1992] :  $\lambda_{ij} + \lambda_{ijk} = \psi_{ij}\phi_k$  (ces  $\psi_{ij}$  et  $\phi_k$  ne sont pas estimés comme ceux du modèle de régression précédent). Les  $\psi_{ij}$  définissent la structure de l'association entre origine sociale et niveau scolaire, elle est supposée identique pour toutes les cohortes. Les  $\phi_k$  eux résument l'intensité de cette association dans les différentes cohortes. Ce modèle, s'il rend compte des données, permet une analyse simple de l'évolution de cette association. Nous discuterons plus en détail l'interprétation des  $\phi_k$  dans la dernière partie, lorsque nous comparerons ce point de vue avec celui que nous allons construire en introduisant la variable latente « potentiel scolaire ».

<sup>5</sup> Traduction de : "no model or measure is the best or the correct way to represent inequality of educational opportunity".

<sup>6</sup> Le logarithme de l'effectif  $m_{ijk}$  des enfants d'origine sociale  $i$ , de niveau scolaire  $j$  dans la cohorte  $k$  est décomposé sous la forme linéaire :  $\ln(m_{ijk}) = \lambda + \lambda_i + \lambda_j + \lambda_k + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + \lambda_{ijk}$  (décomposition classique de l'analyse de variance).

## 2. ÉVOLUTION DES « POTENTIELS SCOLAIRES »

Le paragraphe précédent porte uniquement sur les problèmes généraux soulevés par les comparaisons entre pourcentages ou plus généralement entre tableaux croisés. Les exemples que nous avons utilisés posent d'autres interrogations, comme nous l'avons déjà signalé dans l'introduction.

On peut se demander si les évolutions des taux d'accès à l'enseignement supérieur, pour différentes catégories sociales, permettent de traiter de la démocratisation des formations supérieures. En effet, si on avait utilisé les taux d'accès en sixième pour juger de la démocratisation du collège, on aurait constaté une égalisation à presque 100 % de tous ces taux. Pourtant personne n'aurait pris cet argument comme preuve d'une démocratisation achevée. Les taux d'accès en sixième ne sont plus un indicateur d'inégalité alors qu'ils ont pu l'être dans le passé. La variable « accès aux formations supérieures », comme la variable « accès au collège », a un sens social qui évolue. Sous le même nom on ne mesure pas une variable susceptible de la même interprétation sociologique. Est-il légitime de comparer des taux d'accès aux formations supérieures à des périodes très différentes du point de vue du niveau global de scolarisation ? L'interprétation de l'évolution des taux de scolarisation à différents niveaux scolaires, ne se ramène pas au problème de l'évolution de pourcentages, il y a une difficulté supplémentaire due au fait que les modalités du niveau scolaire sont nominalement les mêmes alors que leurs significations sociales varient dans le temps.

Ce problème de non équivalence d'un même niveau scolaire dans le temps, a été aussi abordé dans le débat alimenté par l'exemple anglais des taux de scolarisation dans le secondaire long. L'évolution globale de la structure éducationnelle influe sur le taux de scolarisation des deux catégories comparées. Pour savoir si les différences observées sont dues à une évolution qui diminue, conserve ou augmente les inégalités, il faut essayer de définir la notion de constance des inégalités dans un système scolaire en mouvement. Le taux de scolarisation à un niveau scolaire donné est un indicateur qui n'a pas la même signification aux différentes dates, il convient donc de ne pas le substantialiser. D. Merllié [1985, p. 647] propose une comparaison parlante :

*on se trouve dans la même situation que si on voulait juger de l'évolution de l'écart des revenus de groupes sociaux en connaissant seulement la proportion de ceux qui dépassent à chaque date un revenu d'un montant constant en francs courants : si l'on passe d'une situation où 1% des ouvriers ont un salaire d'au moins tant par mois contre 37% des cadres à une situation où la même somme est gagnée par 10% des ouvriers et 62% des cadres, comment juger si la tendance a été à un resserrement de la hiérarchie des salaires ? Nul doute que ceux-là mêmes qui concluent à une évolution "incontestable" en matière de scolarisation ne s'interdisent dans ce cas, pourtant formellement identique, à conclure. Pour y voir plus clair – et en admettant qu'on doive se contenter d'un indicateur aussi pauvre que le seul taux de ceux qui atteignent ou dépassent un montant donné –, il faudrait au moins que ce montant ne soit pas "le même" aux deux dates, mais qu'on sache par exemple quelle est, à la deuxième date, la proportion d'ouvriers dont le salaire dépasse celui qui est dépassé par 37 % des cadres. C'est en renonçant à tenir (nominalement) constant l'indicateur retenu qu'on se donnerait un point structurellement fixe de comparaison, permettant de juger si, par rapport à lui, la situation des ouvriers s'est plutôt dégradée ou améliorée.*

Un principe semblable a été utilisé pour la définition d'un seuil de pauvreté dans des pays très différents économiquement, il a produit la notion de seuil relatif de pauvreté.



Cette citation conduit à comparer des taux de scolarisation à des niveaux différents jugés équivalents. Dans son exemple illustratif, D. Merllié [1985, p. 649] ne compare pas les taux d'accès en sixième pour l'année 1962, avec ceux des années 1972-73-74 mais avec les taux d'accès dans les seconds cycles de l'enseignement secondaire<sup>7</sup>. Il justifie cette comparaison de taux d'accès à des niveaux différents, par le fait que ces niveaux sont dans les générations concernées globalement atteints par le même pourcentage d'enfants. Cet argument est aussi utilisé par M. Euriat et C. Thélot [1995, p. 414] pour justifier la comparaison des recrutements de quatre grandes écoles entre 1950 et 1990. Le nombre de places offertes a augmenté mais il représente toujours environ 1,2 % d'une génération. Les auteurs parlent de « comparabilité maximale ». La fréquence globale d'un niveau scolaire est donc prise comme indicateur de sa valeur à la date considérée.

La solution proposée par D. Merllié n'est pas toujours facile à réaliser. Les niveaux scolaires caractéristiques et hiérarchisables n'étant pas très nombreux, il n'est pas toujours facile de trouver pour chaque période considérée un niveau scolaire correspondant au taux d'accès souhaité. La difficulté devient vite insurmontable si on prend en compte plus de deux cohortes ou plusieurs taux d'accès. Pour mettre en œuvre ce principe de comparaison dans des exemples complexes, nous allons essayer de donner à un même niveau scolaire des valeurs différentes dans chaque cohorte, lorsque le taux d'accès change. Cette valeur relative d'un niveau de diplôme dans une cohorte donnée dépend du taux d'accès à ce niveau de la cohorte considérée. Elle est liée au pourcentage des individus qui possèdent dans cette cohorte un diplôme de niveau égal ou supérieur, ceci suppose que les niveaux de diplômes utilisés sont hiérarchisés. Plus ce pourcentage est faible et plus la valeur relative est grande. Cette valeur relative est en fait une valeur relative minimale. Il y a aussi une valeur relative maximale pour ce même niveau de diplôme mais elle correspond à la valeur relative (minimale) du niveau de diplôme immédiatement supérieur. Les niveaux scolaires ne sont plus les courbes de niveau de ces valeurs relatives. D. Merllié [1985] met sur la même courbe de niveau, le niveau d'entrée en sixième en 1962 et l'entrée dans les seconds cycles de l'enseignement secondaire pour les années 1972-1974 sans avoir à évaluer la valeur de cette courbe de niveau. Nous allons essayer de proposer une méthode d'évaluation pour pouvoir traiter des exemples complexes où l'on ne peut pas systématiquement se situer sur des courbes de niveau. Ces valeurs relatives définiront une variable latente que nous proposons d'appeler « potentiel scolaire », car elle permet de comparer les potentiels donnés par les différents diplômes en prenant en compte le niveau concurrentiel de la cohorte. À chaque niveau de diplôme correspond un intervalle de potentiels scolaires, celui défini par la valeur relative (minimale) de ce niveau de diplôme et celle du niveau de diplôme immédiatement supérieur. La borne inférieure est liée au pourcentage des individus qui possèdent un niveau de diplôme au moins égal et la borne supérieure au pourcentage de ceux qui possèdent un niveau strictement supérieur.

La variable latente « potentiel scolaire » définit un indice de réussite scolaire, il ne peut pas être considéré comme un indice de valeur d'échange sur le marché du travail. Cet indice de réussite scolaire concerne l'une des trois missions principales<sup>8</sup> que L. Chauvel [2000, p. 341] assigne à l'école : le classement de tous les individus sur une échelle de qualifications scolaires dans l'intention de les positionner ultérieurement dans

<sup>7</sup> Les taux d'accès en sixième en 1972-73-74 sont de plus peu pertinents pour la recherche d'un indice global d'inégalité, car tous proches de 100 %. Bien entendu, l'étude des non scolarisés en sixième montrerait sûrement une répartition sociale très inégale.

<sup>8</sup> Les deux autres missions sont : la diffusion du savoir ; l'apprentissage de la citoyenneté et de la vie en commun.

la structure sociale. La valeur d'échange sur le marché du travail dépend, elle, de la structure sociale, donc de son évolution. Dans deux cohortes suffisamment éloignées, deux diplômes ayant le même « potentiel scolaire » peuvent conduire à des positionnements moyens dans la structure sociale très différents. Ce problème nous semble plus simple à considérer que celui de la rentabilité sur le marché de l'emploi d'un diplôme donné [C. Baudelot, M. Glaude, 1989]. La réponse à cette dernière question est influencée par deux évolutions, celle de la structure éducationnelle et celle de la structure sociale. Ces évolutions ne sont pas réductibles l'une à l'autre, ce qui peut créer parfois des effets d'apparence paradoxale. Par exemple, dans les sociétés méritocratiques, R. Boudon [1973] montre la possibilité d'une faible liaison entre le statut social relatif des enfants par rapport au père et le niveau d'instruction relatif. La variable latente « potentiel scolaire » peut permettre de simplifier l'étude de l'interaction complexe entre les structures éducationnelle et sociale. D'ailleurs C. Baudelot et M. Glaude [1989] utilisent aussi « un niveau d'éducation relatif », ils s'intéressent au rendement sur le marché de l'emploi du rang de sortie à l'intérieur de chaque génération.

La seule contrainte qui est imposée à la valeur relative (minimale) d'un niveau de diplôme dans une cohorte donnée est d'être une fonction décroissante du pourcentage des individus de cette cohorte qui atteignent un niveau supérieur ou égal. On pourrait donc prendre comme valeur relative le pourcentage des individus qui n'atteignent pas ce niveau de diplôme. On définirait ainsi une variable latente dont la distribution serait uniforme entre 0 % et 100 % (histogramme plat). Ce type de distribution n'est pratiquement jamais trouvé sur les variables observées en sciences humaines. Le choix du type de distribution d'une variable latente continue est classiquement la loi normale. Dans notre cas la variable latente est définie sur l'ensemble de la population, les théorèmes limites centraux du calcul des probabilités peuvent en partie justifier le choix de la loi normale. Cette loi peut être vue comme le mélange des distributions du « potentiel scolaire » sur les différentes catégories sociales. Les répartitions des diplômes dans ces catégories présentent des dissymétries opposées [C. Thelot, L.-A. Vallet, 2000, p. 11]. Cette propriété n'est pas en contradiction avec la symétrie<sup>9</sup> de la loi proposée pour l'ensemble de la cohorte.

Nous allons appliquer la notion de variable latente : « potentiel scolaire » à un exemple qui porte sur 19 générations triennales (entre 1920 et 1976), 11 origines sociales et 7 niveaux de diplôme. Ces données m'ont été aimablement communiquées par L.-A. Vallet. Elles ont été construites à partir des enquêtes Emploi qui se sont déroulées entre 1984 et 2001. Dans chaque génération triennale les échantillons sont suffisamment grands<sup>10</sup> pour que l'on puisse, en première approximation, se limiter à utiliser la statistique descriptive. La construction des données à partir des enquêtes Emploi est expliquée et analysée dans le rapport de L.-A. Vallet et M. Selz [2005]. Le détail du découpage en 11 origines sociales et en 7 niveaux de diplôme est donné en annexe.

C'est à partir des 7 niveaux de diplôme que nous allons construire notre variable latente « potentiel scolaire ». Pour chaque cohorte elle doit fournir des valeurs relatives, qui positionnent les niveaux de diplôme les uns par rapport aux autres. Ces valeurs seront comparables d'une cohorte à l'autre si nous utilisons une même standardisation

<sup>9</sup> Un autre choix que la loi normale pourrait être fait, en particulier des lois dissymétriques : bêta, gamma, log-normale par exemple. Mais le choix d'une densité de référence, d'un type de dissymétrie, pose problème. Il est difficile de trouver une justification expérimentale pour une variable non observable.

<sup>10</sup> Les deux générations triennales extrêmes, 1920-1922 et 1974-1976, sont les moins bien représentées, 4945 et 3750 individus. Pour les autres les effectifs varient de 9764 à 47334.

de la variable latente dans chaque cohorte. Comme la plupart des variables latentes, le « potentiel scolaire » n'a pas d'origine, ni d'unité « naturelles ». On peut toujours choisir ces dernières de telle sorte que la moyenne soit nulle et la variance égale à 1. Ceci ne change pas les résultats des comparaisons. Nous avons fait le choix classique d'une variable latente normale. La valeur relative (minimale) d'un niveau de diplôme est alors la valeur pour laquelle la fonction de répartition de la loi normale centrée réduite est égale à la fréquence des individus qui ont un niveau strictement inférieur.

Dans une cohorte, le « potentiel scolaire » de chaque niveau de diplôme est un intervalle qui dépend de deux fréquences, celles des individus qui ont un niveau de diplôme strictement inférieur d'une part et strictement supérieur d'autre part. Ceci suppose que l'on puisse ordonner les niveaux de diplôme sélectionnés. Les grandes étapes du cursus scolaire retenues dans les enquêtes Emploi ne posent pas trop de problème (cf. l'Annexe). Nous avons simplement réuni le niveau 3 (BEPC sans diplôme technique ou professionnel) et le niveau 4 (CAP, BEP avec ou sans BEPC), car ils nous semblaient difficilement ordonnables. Comme tous les niveaux de diplôme, ce niveau 3-4 ne donne pas un « potentiel scolaire » unique mais un potentiel variant entre un minimum et un maximum. Le minimum est défini en utilisant la fréquence des individus n'ayant pas ce niveau. Par exemple dans la génération 1920-22 (notée g21) la fréquence des individus n'ayant pas atteint le niveau 3-4 est égale à 73.54 %, ce qui donne un potentiel scolaire minimum pour ce niveau égal à  $0.63^{11}$ . Il n'est plus que de -1.20 pour la génération 1974-1976<sup>12</sup>.

Comme le montre le graphique 1, le niveau 3-4 (CAP-BEP-BEPC) est passé d'un potentiel globalement positif, plus de la moitié des individus ne l'atteignent pas<sup>13</sup>, à un potentiel globalement négatif, plus de la moitié des individus le dépassent.

Les lignes horizontales de ce graphique définissent les courbes de niveau du « potentiel scolaire ». Dans la génération 1974-1976 (g75), s'arrêter au niveau baccalauréat donne un potentiel scolaire moyen semblable<sup>14</sup> à l'arrêt au certificat dans la génération 1932-1934 (g33). Il y a cependant plus d'individus qui s'arrêtaient au certificat dans cette dernière génération qu'au baccalauréat dans celle de 1974-1976.

Remarquons que la plage de « potentiel scolaire » trouvée pour un niveau de diplôme ne dépend pas des autres niveaux pris en compte. On peut donc comparer des générations pour lesquelles les niveaux de diplôme sélectionnés ne sont pas les mêmes. Par exemple le niveau certificat peut très bien disparaître pour les dernières générations<sup>15</sup>. Par contre il devient urgent d'introduire d'autres niveaux de diplôme pour stratifier la plage grandissante des 2<sup>e</sup> et 3<sup>e</sup> cycles du supérieur. Ceci précisera la variable latente mais n'affectera pas les potentiels des niveaux déjà utilisés. La variable latente « potentiel scolaire » peut permettre des comparaisons entre des structures de diplômes différentes, entre pays ou générations éloignées par exemple.

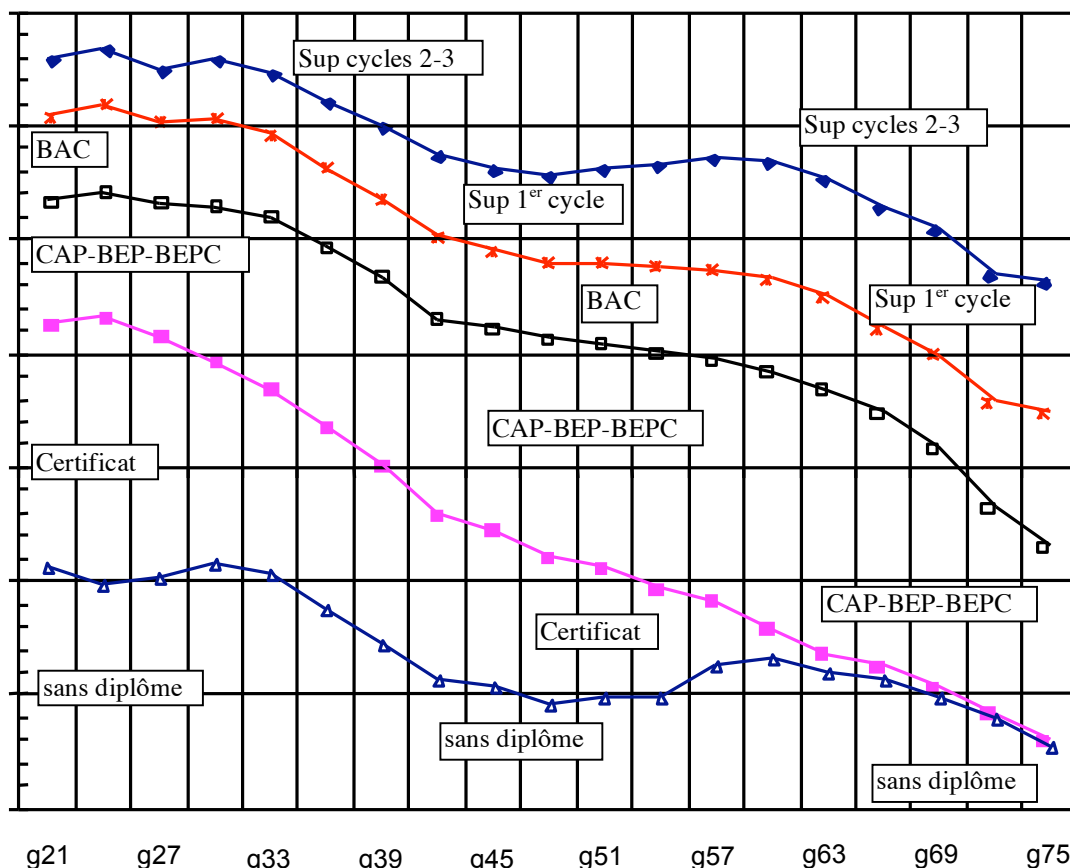
<sup>11</sup>  $0.63 = \Phi^{-1}(73.54 \%)$ ,  $\Phi$  étant la fonction de répartition de la loi  $N(0,1)$ .

<sup>12</sup>  $-1.20 = \Phi^{-1}(11.59 \%)$  car pour cette génération il n'y a plus que 11.59 % des individus qui n'ont pas atteint le niveau 3-4.

<sup>13</sup> La médiane générale correspond à la valeur 0 (c'est aussi la moyenne générale) puisque pour chaque génération le « potentiel scolaire » suit la loi normale centrée réduite.

<sup>14</sup> Ces moyennes sont un peu inférieures à 0,0. Ce sont les moyennes de la loi normale centrée réduite sur les intervalles de potentiel du baccalauréat en g75 et du certificat en g33.

<sup>15</sup> Le certificat d'études primaires a été transformé en une épreuve pour adultes en 1971, ce qui explique la remontée de la courbe inférieure à la génération 1956-1958 (g57). Il a été aboli en 1989 [C. Verrier, 2001].



GRAPHIQUE 1. Évolution des « potentiels scolaires » des diplômes

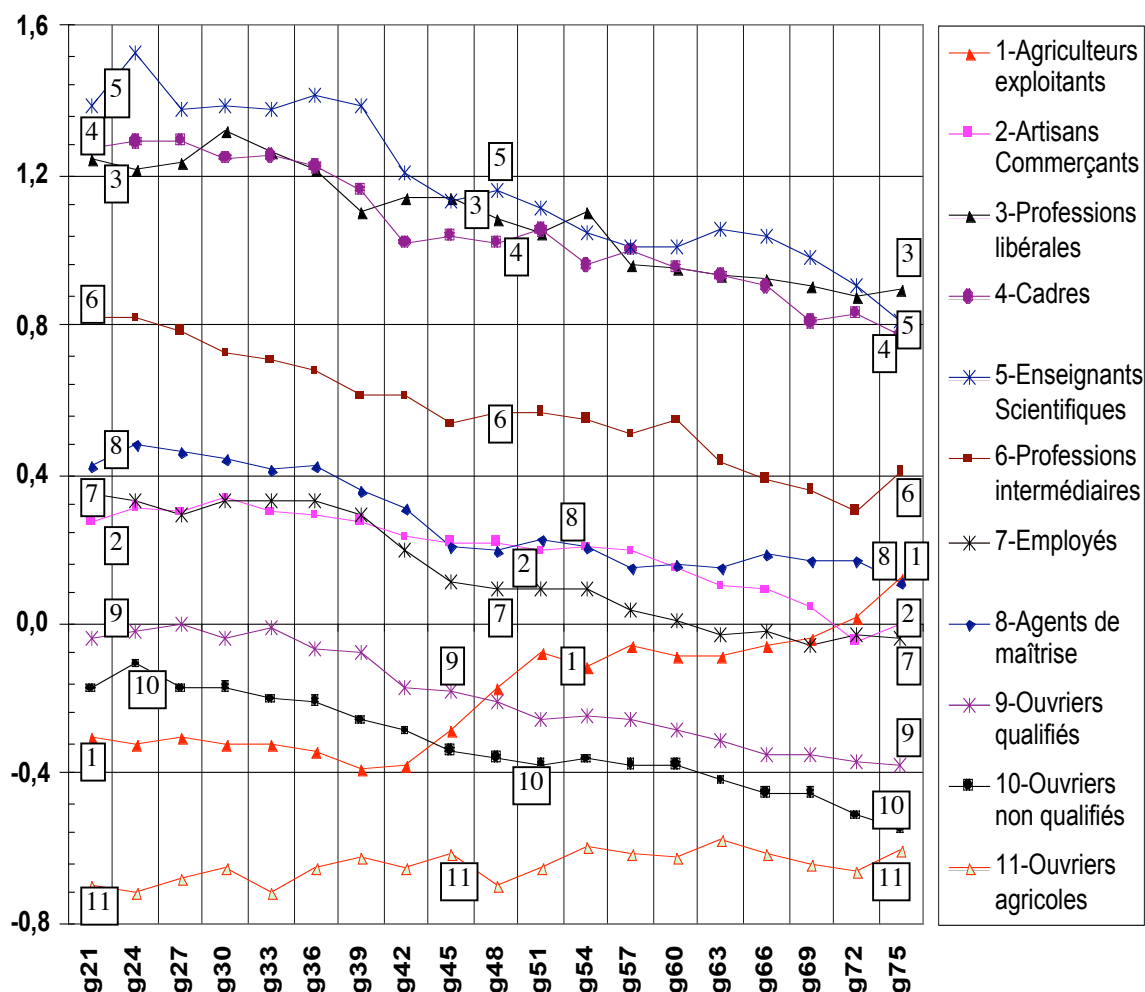
Lecture : pour la génération 1926-1928 (g27) le certificat donne un potentiel scolaire variant entre -0,5 et 0,6. Dans la génération g27 la fréquence des individus qui se sont arrêtés au certificat est égale à la probabilité de l'intervalle  $]-0,5 ; 0,6]$  pour une loi normale centrée réduite.

La décroissance globale des courbes du graphique correspond à l'élévation du niveau d'éducation. La diffusion des diplômes confère automatiquement à chacun d'eux une valeur relative en diminution. Cette décroissance à peu près continue est marquée par deux périodes d'accélération, la plus marquée pour les générations autour des années 1934-1943, l'autre autour des années 1964-1973. L. Chauvel [1998, p. 8] trouve aussi deux périodes d'accélération semblables en étudiant l'évolution générationnelle de l'âge médian de fin d'études initiales. La première accélération semble liée à l'après guerre et à la réforme Carcopino de 1941 qui rattache les écoles primaires supérieures au second degré [A. Prost, 1997, p. 89-92]. La deuxième accélération semble liée à l'après Mai 68, à l'unification des collèges qui s'achève par la réforme Haby de 1975 dite du « collège unique » [A. Prost, 1997, p. 101-103] et à la réforme Chevènement de 1985 avec la création des baccalauréats professionnels associée à l'objectif de 80 % d'une classe d'âge au niveau du baccalauréat [A. Prost, 1997, p. 205-210].

### 3. ORIGINE SOCIALE ET « POTENTIEL SCOLAIRE »

Nous allons maintenant regarder comment évolue la distribution du « potentiel scolaire » dans les différentes catégories sociales retenues. Commençons par regarder l'évolution des positions des origines sociales les unes par rapport aux autres à partir d'un indice de tendance centrale. Le graphique 2 correspond à la médiane, le suivant utilise la moyenne<sup>16</sup>.

Les deux indices de tendance centrale, médiane et moyenne, donnent des graphiques semblables qui conduisent aux mêmes conclusions.

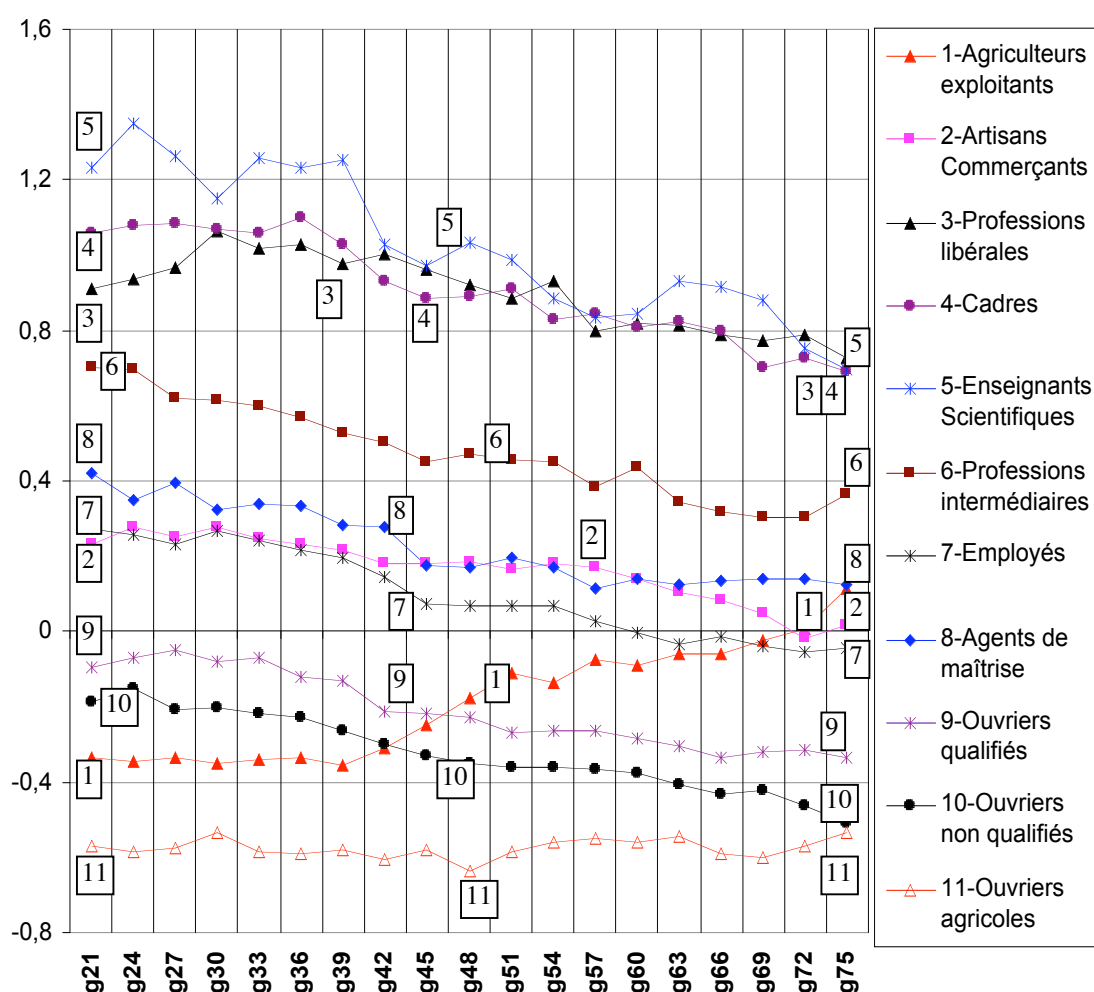


GRAPHIQUE 2. Évolution des potentiels médians dans 11 catégories sociales

Lecture : pour la génération 1926-1928 (g27) le potentiel scolaire médian des ouvriers qualifiés (courbe 9) est égal à 0,0.

<sup>16</sup> Pour chaque génération, les 6 niveaux de diplômes retenus permettent de découper le « potentiel scolaire » en 6 intervalles consécutifs. Les données nous permettent d'avoir la répartition sur ces intervalles des enfants de chaque catégorie. Nous calculons la médiane de ces 19x11 histogrammes. Le calcul des moyennes est fait en utilisant pour chaque intervalle de potentiel la moyenne de la loi normale réduite sur cet intervalle.

Deux courbes ont une évolution très différente des autres, celle des enfants d'agriculteurs exploitants (1) et celle des enfants d'ouvriers agricoles (11). Les potentiels médians et moyens de ces deux catégories sont les seuls à ne pas présenter une tendance décroissante. Les enfants d'agriculteurs sont partis d'un « potentiel scolaire » médian ou moyen inférieur à celui des enfants d'ouvriers pour atteindre celui des enfants d'agents de maîtrise, d'artisans, de commerçants ou d'employés. Les enfants d'ouvriers agricoles qui étaient de très loin les moins diplômés ont rattrapé les enfants d'ouvriers. Le monde agricole paraît avoir trouvé une position, par rapport au système scolaire, semblable à celle des autres catégories. La métamorphose quantitative<sup>17</sup> et qualitative du monde agricole a entraîné une diminution du potentiel médian et moyen des autres catégories. Le non investissement dans le système scolaire des enfants d'agriculteurs et d'ouvriers agricoles valorisait mécaniquement les potentiels scolaires des diplômes des autres catégories.



GRAPHIQUE 3. Évolution des potentiels moyens dans 11 catégories sociales

Lecture : pour la génération 1926-1928 (g27) le potentiel scolaire moyen des agents de maîtrise (courbe 8) est égal à 0,4.

<sup>17</sup> Les enfants d'agriculteurs et d'ouvriers agricoles représentent 32,9% de la génération 1920-22 et seulement 6,1% de la génération 1974-76.

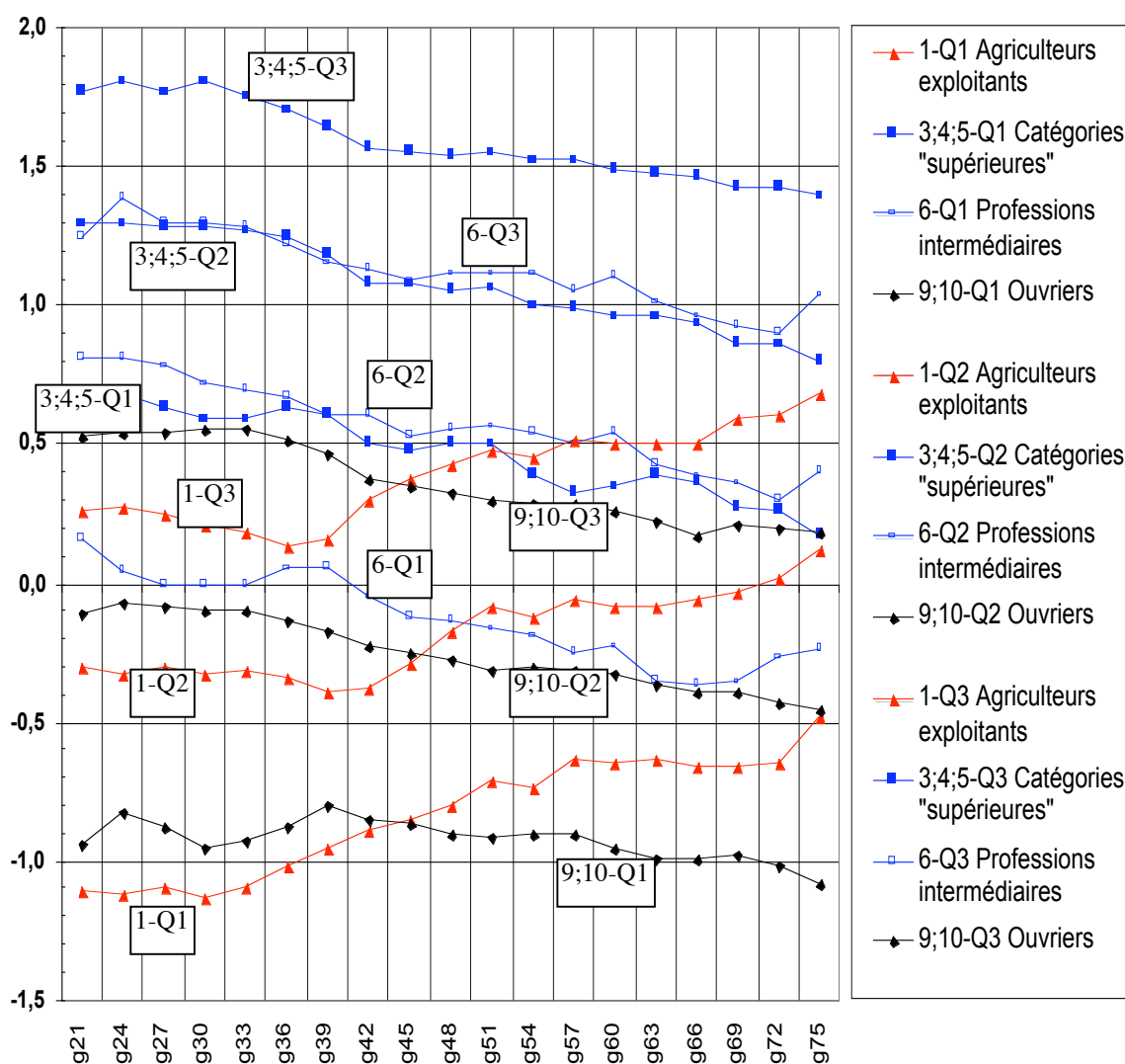
Le deuxième fait marquant des Graphiques 2 et 3, est l'extraordinaire parallélisme de l'évolution des potentiels médians et moyens des catégories sociales non agricoles. On distingue en fait quatre ensembles homogènes : le groupe des catégories « supérieures » (3-Professions libérales<sup>18</sup>, 4-Cadres, 5-Enseignants Scientifiques), celui des professions intermédiaires (6), le groupe des catégories « moyennes » (2-Artisans Commerçants, 7-Employés, 8-Agents de maîtrise) et celui des ouvriers (9 et 10). La diminution du potentiel médian est de l'ordre de 0,3 à 0,5 alors que pour les catégories agricoles c'est une augmentation de 0,1 à 0,5 qui est observée. Pour juger de l'évolution de l'inégalité scolaire, il faudra prendre en compte cette opposition entre le positionnement très stable des catégories non agricoles dans la structure scolaire relative décrite par le « potentiel scolaire » et l'évolution différente des potentiels des enfants d'agriculteurs et d'ouvriers agricoles.

La variable latente « potentiel scolaire » permet aussi d'analyser l'évolution de l'homogénéité de chaque catégorie par rapport à sa position relative dans le système scolaire. L'indice de dispersion associé à la médiane est généralement la longueur de l'intervalle interquartiles. Le calcul de cet indice sur les 11 catégories, pour les 19 générations montre que les différences entre catégories ont une évolution qui tend à les rapprocher. Dans les premières générations, la variable « potentiel scolaire » était plus hétérogène chez les enfants d'ouvriers que chez les enfants de cadres ou de catégories intermédiaires, ce n'est plus le cas dans les dernières générations. L'ensemble des catégories a évolué vers un intervalle interquartile de longueur assez proche. La dispersion du « potentiel scolaire » est devenue semblable pour toutes les catégories. La longueur de l'intervalle interquartile, de l'ordre de  $1,22 \pm 0,15$ , reste cependant forte, car la différence entre les potentiels scolaires médians extrêmes n'est que de l'ordre de 1,5 pour les dernières générations (cf. Graphique 2).

Sur le Graphique 4 nous avons tracé l'évolution des trois quartiles de quatre catégories sociales (Q1 = quartile inférieur, Q2 = médiane et Q3 = quartile supérieur).

Pour les catégories « ouvriers » (9 + 10) et les catégories « supérieures » (3 + 4 + 5) les intervalles interquartiles ne se recouvrent jamais. Le quartile inférieur des catégories « supérieures » représente un potentiel scolaire qui est dépassé par moins de 25 % des enfants d'ouvriers alors que, par construction, il est dépassé par 75 % des enfants des catégories « supérieures ». Pour les enfants des professions intermédiaires le quartile supérieur est très proche de la médiane des catégories « supérieures », seulement 25 % de ces enfants accèdent à un potentiel scolaire supérieur à cette médiane. Ces inégalités apparaissent extrêmement stables, les courbes des quartiles évoluent parallèlement. Sur l'ensemble des 11 catégories sociales, on retrouve aussi une évolution semblable à celle de la médiane pour le quartile inférieur et pour le quartile supérieur. Le changement principal est celui du monde agricole.

<sup>18</sup> Pour les premières générations étudiées, les professions libérales se différencient un peu des catégories 4 et 5 quand on utilise les moyennes. Les enfants sans diplôme y sont plus fréquents.



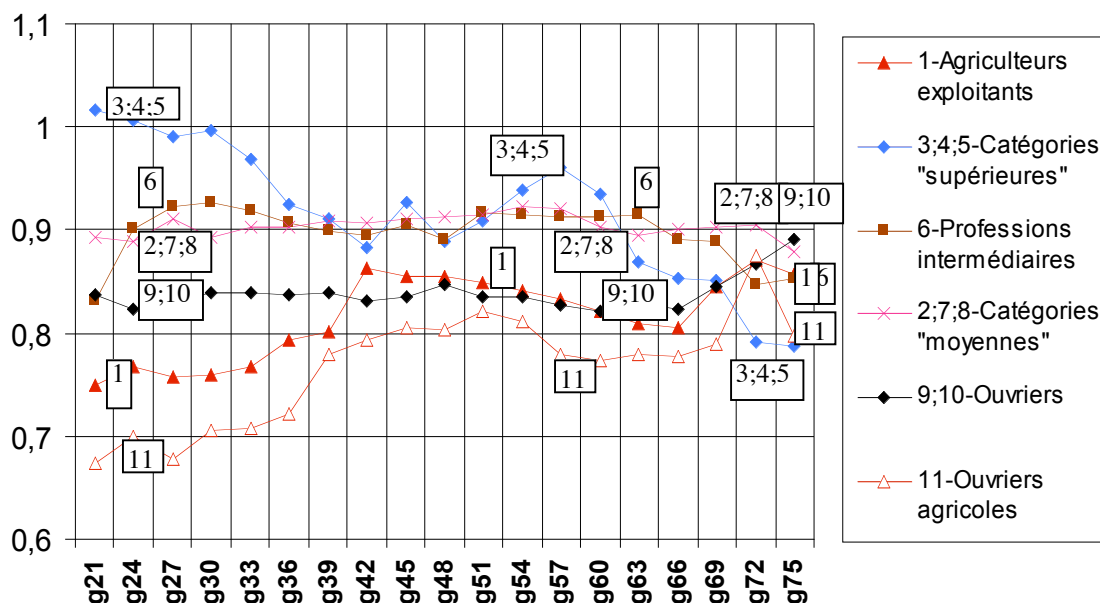
GRAPHIQUE 4. Évolution des quartiles du « potentiel scolaire » dans 4 catégories sociales

Lecture : pour la génération 1926-1928 (g27) les quartiles du potentiel scolaire des agriculteurs exploitants sont égaux à -1,1 ; -0,3 et 0,3 (courbes 1-Q1 ; 1-Q2 et 1-Q3).

L'étude de l'évolution de l'homogénéité de la variable « potentiel scolaire » dans les différentes catégories sociales peut aussi se faire en regardant l'écart-type.

Comme précédemment on trouve que les valeurs de cet indice de dispersion se rapprochent. Dans les premières générations, les enfants du monde agricole sont plus homogènes par rapport au système scolaire que les enfants des catégories « supérieures ». Ensuite les hétérogénéités des potentiels scolaires au sein des différentes catégories deviennent semblables. Les différentes catégories sociales retenues conduisent à des potentiels scolaires moyens toujours bien différenciés, avec des hétérogénéités internes qui se sont rapprochées mais qui restent importantes.





GRAPHIQUE 5. Évolution des écart-types du « potentiel scolaire »

Lecture : pour la génération 1938-1940 (g39) l'écart-type du potentiel scolaire des agriculteurs exploitants (courbe 1) est égal à 0,8.

#### 4. DÉMOCRATISATION ?

La démocratisation de l'enseignement fait l'objet de nombreuses études, pourtant comme le note P. Merle [2002, p. 635] « il n'existe en fait ni définitions incontestées ni modalités de mesure codifiées et admises par tous ». Il en est de même du découpage introduit par A. Prost [1986, p. 12] et souvent réutilisé : la démocratisation « quantitative » et la démocratisation « qualitative ». Ce découpage repose sur le besoin d'essayer de mettre à part l'effet de l'allongement général de la scolarité. Cette « massification » peut se traduire par une « translation » des inégalités, on a alors une démocratisation simplement « quantitative ». Comme la démocratisation de la voiture, la « massification » n'est pas un indice de diminution des inégalités. Tout le problème est la traduction de cette « translation » des inégalités. Dans la modélisation log-linéaire elle correspond à l'absence d'interaction d'ordre trois : les rapports des chances sont alors dans les mêmes rapports d'une cohorte à l'autre. Dans le modèle log-multiplicatif de Yu Xie (voir la première partie) cela se traduit par l'égalité des  $\phi_k$ .

Avec notre point de vue la démocratisation « quantitative » est prise en compte par la définition de la variable latente « potentiel scolaire ». Étudier la valeur relative des diplômes est une manière d'annuler l'effet de la hausse générale du niveau de formation. Cette hausse conduit à une décroissance au cours du temps des valeurs relatives d'un même niveau de scolarité. Le graphique 1 que nous avons déjà analysé donne à voir une « massification » générale et continue avec deux périodes d'accélération. La démocratisation « quantitative » est incontestable et il n'y a pas de discussion sur le constat même s'il peut y en avoir sur la définition de cette démocratisation.

Le travail sur les potentiels scolaires est une manière d'éliminer la démocratisation « quantitative ». La démocratisation « qualitative » se traduit alors dans l'évolution des caractéristiques des potentiels scolaires des différentes origines sociales.

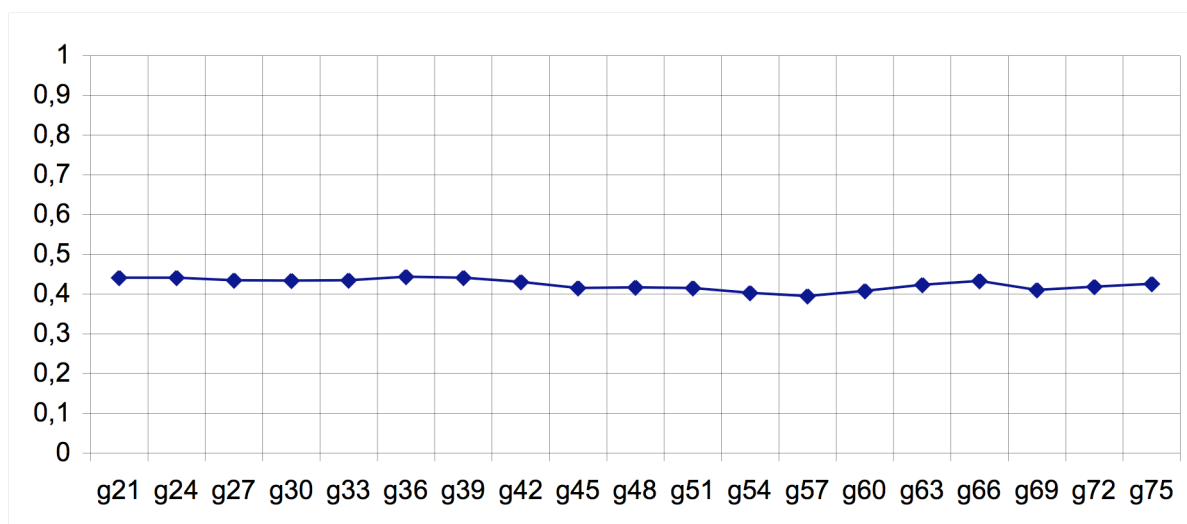
Les évolutions des médianes et des moyennes (Graphiques 2 et 3) mais aussi des quartiles définissent des courbes extraordinairement parallèles pour les catégories non agricoles. Seules les courbes des catégories agriculteurs exploitants et ouvriers agricoles se distinguent. Les enfants du monde agricole étaient les moins scolarisés dans les premières générations étudiées, ils se sont ensuite de plus en plus investis dans le système scolaire, comparativement aux autres catégories. Cet investissement et la forte diminution en effectif de ces catégories expliquent la décroissance observée pour les catégories non agricoles. D'autant plus que dans les premières générations le rapport au système scolaire du monde agricole était plutôt homogène (cf. Graphique 5).

Les enfants des ouvriers agricoles rejoignent simplement ceux des ouvriers non qualifiés (cf. les Graphiques 2 et 3). Les courbes de tendance centrale des enfants d'agriculteurs exploitants (Graphiques 2 et 3) montrent une croissance accélérée à partir des générations d'après guerre<sup>19</sup>, elles rejoignent celles des agents de maîtrise, artisans, commerçants et employés. Pour les ouvriers agricoles, on peut sans problème parler d'une démocratisation « qualitative ». Il semble en être de même pour les agriculteurs exploitants, mais ce n'est pas aussi simple, une part de cette mutation scolaire peut être liée à la transformation du métier d'agriculteur exploitant. Les caractéristiques de cette catégorie sont-elles suffisamment comparables entre les générations 1920 et 1976 pour que l'on puisse penser qu'elle occupe la même position dans l'espace social ? Si la réponse est non, l'évolution positive du potentiel scolaire de cette catégorie ne peut pas complètement passer pour de la démocratisation « qualitative ». Cette remarque nous montre que la substantialisation d'une catégorie sociale peut poser problème dans l'interprétation d'une évolution. Comme nous l'avons vu dans l'introduction la substantialisation d'une variable porte aussi bien sur la population concernée que sur la « mesure » effectuée. En sciences humaines les variables souffrent souvent de ce manque de stabilité dans le temps, ce qui complique les comparaisons diachroniques. L'interprétation des résultats des méthodes statistiques doit alors se faire avec précaution, il est facile de se faire piéger par ces variables dont le sens évolue alors que la définition ne bouge pas.

Trouver un bon indice de démocratisation « qualitative » n'est déjà pas une chose simple avec des variables ordinaires (cf. le débat analysé dans la partie 1), alors dans le cas de variables « molles »... En utilisant la variable « potentiel scolaire » nous pouvons nous demander si l'influence des catégories sociales sur cette variable diminue ou augmente au cours du temps. Un résumé classique de cette influence peut être donné par la valeur du rapport de corrélation<sup>20</sup>. Le Graphique 6 montre une constance de cet indice. Le déterminisme social sur le « potentiel scolaire » oscille entre 0.40 et 0.44. Il y a cependant eu des changements : les agriculteurs exploitants ont rejoint les catégories « moyennes », les ouvriers agricoles ont rejoint les ouvriers, les variances intra-catégories se sont homogénéisées et les poids des différentes catégories ont changé. Cette liste de changements renvoie principalement à l'évolution du monde agricole.

<sup>19</sup> Les collèges et lycées agricoles ont été créés en 1960 [C. Verrier, 2001].

<sup>20</sup> Le carré du rapport de corrélation étant le rapport entre la variance inter et la variance totale, il est ici égal à la variance inter puisque la variance de la variable latente est normalisée à 1.



GRAPHIQUE 6. Évolution du rapport de corrélation

L.-A. Vallet et M. Selz [2005, p. 19-23 ; 2006, p. 6] retrouvent cette importance du monde agricole en analysant les mêmes données avec le modèle log-multiplicatif de variation uniforme. Ce modèle permet une mesure de l'intensité de l'association statistique entre origine sociale et diplôme, elle diminue nettement entre 1920 et 1976 lorsqu'on prend en compte l'ensemble des catégories, mais diminue beaucoup moins lorsqu'on enlève les agriculteurs exploitants et les ouvriers agricoles. On retrouve cette différence d'évolution en utilisant comme indice de liaison, le coefficient de contingence de Cramer.

La contradiction entre la diminution de l'indice du modèle log-multiplicatif de variation uniforme et la constance du rapport de corrélation précédent n'a rien d'étonnant, les indices statistiques globaux de l'inégalité posent d'énormes problèmes de choix<sup>21</sup> et d'interprétation. Pour bien analyser ces indices il faut souvent les décomposer en informations plus fines, ce qui élimine en partie le côté synthétique.

La notion de démocratisation « qualitative » est suffisamment difficile à cerner pour que sa quantification conduise à des solutions différentes pouvant se contredire. Dans notre exemple la démocratisation « qualitative » semble principalement toucher les enfants d'ouvriers agricoles et ceux des agriculteurs exploitants. Ce n'est pas la même chose que de dire qu'il y a eu démocratisation « qualitative » d'un point de vue global. Ce type de démocratisation est une construction liée à l'analyse de l'évolution du système scolaire, en particulier des réformes qui l'ont ponctuée. La plupart de ces réformes ne visaient pas spécialement les enfants du monde agricole, il apparaît donc difficile de penser que l'évolution structurelle du système scolaire a été une source de démocratisation « qualitative ».

<sup>21</sup> Les différentes solutions proposées dans le débat lancé par J.-C. Combessie n'aboutissent pas aux mêmes conclusions. Il existe rarement un indice vérifiant toutes les propriétés dont l'utilisateur aimerait le voir paré [Barbut, 1985 ; Morel, 2003].

## BIBLIOGRAPHIE

- BARBUT M., « Sur quelques propriétés élémentaires des fonctions de concentration de C. Gini », *Mathématiques et Sciences humaines* 88, 1984, p. 5-19.
- BARBUT M., « Sur les indicateurs de l'inégalité : croissance logistique et mesure de l'inégalité, et de quelques effets 'paradoxaux' dans la comparaison des inégalités », *Mathématiques et Sciences humaines* 90, 1985, p. 5-17.
- BAUDELOT C., GLAUDE M., « Les diplômes se dévaluent-ils en se multipliant ? », *Economie et statistique* 225, 1989, p. 3-16.
- BOUDON R., *L'inégalité des chances. La mobilité sociale dans les sociétés industrielles*, Paris, Armand Colin, 1973.
- CHAUVEL L., « La seconde explosion scolaire : diffusion des diplômes, structure sociale et valeur des titres », *Revue de l'OFCE* 66, 1998, p. 5-36.
- CHAUVEL L., « Valorisation et dévalorisation sociale des titres scolaires : une comparaison France – Etats-Unis », A. Van Zanten (ed.), *L'école Texte imprimé : l'état des savoirs*, Paris, la Découverte, 2000, p. 341-352.
- COMBESSIE J.-C., « L'évolution comparée des inégalités : problèmes statistiques », *Revue française de sociologie* 25 (2), 1984, p. 233-254.
- COMBESSIE J.-C., « Trente ans de comparaison des inégalités des chances : quand la méthode retenue conditionne la conclusion », *Courrier des statistiques* 112, décembre, 2004, p. 37-54.
- COUR DES COMPTES, *La gestion du système éducatif*, Rapport au président de la république, Paris, Cour des comptes, 2003.
- DURU-BELLAT M., KIEFFER A., « La démocratisation de l'enseignement en France : polémiques autour d'une question d'actualité », *Population* 55 (1), 2000, p. 51-80.
- EURIAT M., THELOT C., « Le recrutement social de l'élite scolaire en France. Évolution des inégalités de 1950 à 1990 », *Revue française de sociologie* 36 (4), 1995, p. 403-438.
- FLORENS J.-P., « Inégalité et dépendance statistique », *Revue française de sociologie* 25 (2), 1984, p. 255-263.
- GOODMAN L. A., HOUT M., "Statistical methods and graphical displays for analyzing how the association between two qualitative variables differs among countries, among groups, or over time: a modified regression-type approach", *Sociological Methodology* 28, 1998, p. 175-230.
- GRÉMY J.-P., « Sur les différences entre pourcentages et leur interprétation », *Revue française de sociologie* 25 (3), 1984, p. 396-420.
- MARE R. D., "Change and stability in educational stratification", *American Sociological Review* 46, 1981, p. 72-87.
- MERLE P., « Démocratisation ou accroissement des inégalités scolaires ? L'exemple de l'évolution de la durée des études en France (1988-1998) », *Population-F* 57 (4-5), 2002, p. 633-660.
- MERLLIÉ D., « Analyses de l'interaction entre variables. Problème statistique ou sociologique ? », *Revue française de sociologie* 26 (4), 1985, p. 629-652.
- MOREL G., « Modéliser l'évolution de l'inégalité des taux de scolarisation », *Mathématiques et Sciences humaines* 164, 2003, p. 5-27.
- PRÉVOT J., « À propos d'indices et de comparaisons de proportions », *Revue française de sociologie* 26 (4), 1985, p. 601-628.
- PROST A., *L'enseignement s'est-il démocratisé ?*, Paris, Presses Universitaires de France, 1986.
- PROST A., *Education, société et politiques. Une histoire de l'enseignement de 1945 à nos jours*, Paris, Seuil, 1997.
- SELZ M., VALLET L.-A., « La démocratisation de l'enseignement et son paradoxe apparent », *Données sociales – La société française*, Paris, INSEE, 2006, p. 101-107.

TÉROUANNE E., « Distorsion entre deux distributions d'une variable nominale », *Mathématiques, Informatiques et Sciences humaines* 131, 1995, p. 29-38.

THELOT C., VALLET L.-A., « La réduction des inégalités sociales devant l'école depuis le début du siècle », *Économie et statistique* 334, 2000, p. 3-32.

VALLET L.-A., « L'évolution de l'inégalité des chances devant l'enseignement. Un point de vue de modélisation statistique », *Revue française de sociologie*, 29 (3), 1988, p. 395-423.

VALLET L.-A., SELZ M., *Evolutions récentes et état actuel de l'équité du système éducatif en France*, Paris, Rapport de recherche pour le Commissariat général du plan et la Direction de l'évaluation et de la prospective, 2005.

VALLET L.-A., SELZ M., « Évolution historique de l'inégalité des chances devant l'école : des méthodes et des résultats revisités », *Éducation & formation* 74, 2007, p. 65-74.

VERRIER C., *Chronologie de l'enseignement et de l'éducation en France*, Paris, Economica, 2001.

WESTERGAARD J., LITTLE A., "Educational opportunity and social selection in England and Wales. Trends and policy implications", *Social Objectives in Educational Planning*, Paris, OCDE, 1967.

XIE Y., "The log-multiplicative layer effect model for comparing mobility tables", *American Sociological Review* 57(3), 1992, p. 380-395.

## ANNEXE

### NOMENCLATURE D'ORIGINE SOCIALE :

1. Agriculteurs exploitants (PCS 10) ;
2. Artisans ; Commerçants et assimilés (PCS 21 et 22) ;
3. Chefs d'entreprise de 10 salariés ou plus ; Professions libérales (PCS 23 et 31) ;
4. Cadres de la fonction publique ; Professions de l'information, des arts et des spectacles ; Cadres administratifs et commerciaux d'entreprises ; Ingénieurs et cadres techniques d'entreprises (PCS 33, 35, 37 et 38) ;
5. Professeurs, professions scientifiques ; Instituteurs et assimilés (PCS 34 et 42) ;
6. Professions intermédiaires de la santé et du travail social ; Clergé, religieux ; Professions intermédiaires administratives de la fonction publique ; Professions intermédiaires administratives et commerciales des entreprises ; Techniciens (PCS 43, 44, 45, 46 et 47) ;
7. Employés civils et agents de service de la fonction publique ; Policiers et militaires ; Employés administratifs d'entreprises ; Employés de commerce ; Personnels des services directs aux particuliers (PCS 52, 53, 54, 55 et 56) ;
8. Contremaîtres, agents de maîtrise (PCS 48) ;
9. Ouvriers qualifiés de type industriel ; Ouvriers qualifiés de type artisanal ; Ouvriers qualifiés de la manutention, du magasinage et du transport (PCS 62, 63 et 65) ;
10. Chauffeurs ; Ouvriers non qualifiés de type industriel ; Ouvriers non qualifiés de type artisanal (PCS 64, 67 et 68) ;
11. Ouvriers agricoles (PCS 69).

### NOMENCLATURE EN SEPT POSTES DU DIPLÔME LE PLUS ÉLEVÉ OBTENU :

1. Aucun diplôme ou diplôme non déclaré (codes 71 et blanc) ;
2. Certificat d'études primaires (code 70) ;
3. BEPC sans diplôme technique ou professionnel (code 60) ;
4. CAP, BEP avec ou sans BEPC (codes 50 et 51) ;
5. Baccalauréat ou diplôme de niveau équivalent : baccalauréat général et diplôme technique secondaire ; baccalauréat général seul ; baccalauréat de technicien, baccalauréat technologique, baccalauréat professionnel ou brevet professionnel ; BEI, BEC, BEA (codes 40, 41, 42 et 43) ;
6. Diplôme d'une ou deux années après le baccalauréat : diplôme de 1<sup>er</sup> cycle universitaire ; BTS, DUT ; diplôme paramédical ou social avec ou sans baccalauréat général (codes 30, 31, 32 et 33) ;
7. Diplôme d'au moins trois années après le baccalauréat : diplôme de 2<sup>e</sup> ou 3<sup>e</sup> cycle universitaire ; diplôme de grande école, diplôme d'ingénieur (codes 10 et 11).